

Speech Processing 2015/2016

2nd Test

April 22nd 2016

Please identify this form with your name and student number in the reserved spaces at the beginning and end of the test. The answers to multiple-choice questions will only be accepted if inserted in the appropriate place. Wrong answers to these questions will be penalized. The phonetic symbols should use the SAMPA alphabet (Lisbon accent).

Name:	
Number:	
Group number:	

1. Classify as True (T) or False (F) on the left side of each item.

- (a) In analysis-by-synthesis coders, interpolation of the short-term predictor coefficients should be used both at the transmitter and receiver?
- (b) The first coder using noise shaping techniques was multipulse.
- (c) In CELP coders, the adaptive codeword scheme is just an analysis-by-synthesis approach for deriving the parameters of the long-term predictor.
- (d) In vector quantization of a sequence of N samples, increasing N allows for a better approximation to the minimum distortion bound.
- (e) The excitation of RPE coders is a grid of equally spaced pulses, separated by the pitch period.
- (f) In concatenative synthesis systems, using unit selection, the computation of the concatenation costs involves only spectral (cepstral) parameters.
- (g) Limited vocabulary synthesizers (e.g. for synthesizing credit card numbers) never take prosody into account, just concatenating words.
- (h) Text normalization of money amounts and dates is typically done by data driven approaches.
- (i) In PSOLA synthesis models, the speech signals in the database are first processed to decompose the waveforms into a sequence of pitch-synchronous overlapping windows.
- (j) In a synthesizer, intelligibility may be affected by the target speaking rate.

2. State one advantage and one disadvantage of using two coefficients in the pitch prediction filter, instead of just one (besides the delay).

3. Which of the two coders operates at a lower bit rate, MELP or MBE?

Can they be considered vocoders? Justify briefly.

4. Name two tests that may be used in the assessment of the quality of both coders and synthesizers.

5. How many pitch accent types are necessary to describe tonal events in the ToBI model for EP? _____
6. In Lab 3, when simulating a formant synthesizer for vowels, you were asked to observe the corresponding JND for several parameters. What are the typical values of JND (in Hz) for the F0 of a male voice?

7. Specify the bit allocation for a hypothetical CELP coder, which allows two different operating modes, at 6.6 and 4.4 kbps, changing the fewest possible parameters between the two modes.

Suggestion: An efficient way to encode the lag of the pitch predictor in multipulse and CELP-like coders is to do it differentially every other frame, for instance, allocating just 2 bits, every other frame.

6.6 kbps

Parameter	Updating	No. Bits / Frame	Sub-total
long-term predictor	every		
short-term predictor	every		
excitation	every		

4.4 kbps

Parameter	Updating	No. Bits / Frame	Sub-total
long-term predictor	every		
short-term predictor	every		
excitation	every		

Indicate whether the two operating modes result in different algorithmic delays. (T/F) _____

8. Write the nativised broad phonetic transcription of the following sentence, using only EP phones, as you would hear it in the Portuguese media.

When I find myself in times of trouble, Mother Mary comes to me, speaking words of wisdom, let it be.

9. State the basic rule for pronouncing acronyms of the *siglae* type.

10. Figure 1 shows the waveform corresponding to 3 isolated words: *sobre, sensação, fixador*.

- (a) Which is the order of the words?

- (b) Which phone sequences from which words would you select to concatenate the target word *sensor*, using the fewest sequences. Use # as a symbol for silence. Indicate for each segment the approximate beginning and end (e.g. segment 1 - transcription - 0.15s to 0.35s).

- (c) The prosody generation module computed a target duration of 150ms for the second vowel. Indicate the original durations for the two sub-units and show how you would compute the corresponding modified values.

- (d) Which type of post-processing is necessary to synthesize this word? (for non-Portuguese students)
-

- (e) Can you think of any other word that can be synthesized using units from this waveform ? (for Portuguese students)
-

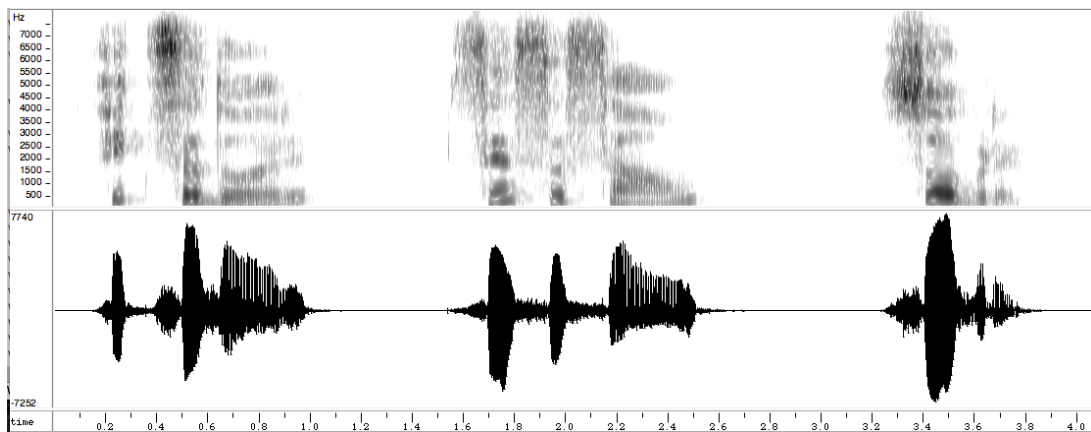


Figure 1

11. In European Portuguese, the pronunciation of grapheme *d* is context independent. In Brazilian Portuguese, there are two alternatives exemplified by the following examples: *adeus*, *admirar*, *adro*, *advogado*, *cidade*, *ciadela*, *doutor*, *dia*, *tardes*.

- (a) What are the two possibilities in terms of manner of articulation?
-

- (b) Using the following syntax

$a \rightarrow b / c \dots d$

describe the simplified rules for grapheme *d*, where *m* and *n* may be graphemes (e.g.: *a*, *b*, etc.), classes of graphemes (vowels, consonants, etc.), the word boundary (*#*), or any grapheme (***). You may use the symbol *0* to mark phonemic nulls, and the symbols *|* and *()* to mark disjunction between several graphemes (e.g.: *a | b | c*). The rules are applied in order, until one matches the context and, in this case, the following rules are not applied. The “simplified” rules do not need to contemplate compound words, or words of foreign origin.

--

1	3.0
2	1.0
3	1.5
4	1.0
5	1.0
6	1.0
7	3.0
8	2.0
9	1.0
10	3.0
11	2.5

Name:	
Number:	
Group number:	